PROVIDING DETERMINISTIC QUALITY OF SERVICE GUARANTEE IN A WIRELESS ENVIRONMENT [1]

Felix K. L. Lee and Mounir Hamdi [2]
Department of Computer Science
The Hong Kong University of Science & Technology
Clear Water Bay, Kowloon
Hong Kong

Abstract - The provision of quality-of-service (QoS) to multimedia applications is very important in future generation wireless networks. In this paper, we propose a *deterministic* QoS guarantees method which is based on a mathematical framework that accuratly characterizes multimedia traffic streams in conjunction with efficient scheduling and dropping algorithms. The uniqueness of this scheme is that it can deterministically provide bounded loss ***and*** bounded delay. We performed a set of evaluation experiments that will demonstrate that our proposed framework can significantly support more connections than a system which do not allow any packet loss under the same required determinitic QoS. We also show that the proposed algorithms are able to out-perform statistical scheduling algorithms adopted in state-of-the-art wireless MAC protocols when the worst-case traffic is being used.

1 Introduction

With rapid advances in wireless transmission technologies, there is a great interest in using wireless networks for the transmission of not only traditional discrete data such as text and still images, but also multimedia traffic such as real-time images, voice, and video [1]. Multimedia applications are different from traditional applications in that they require QoS guarantees in terms of delay, delay variation, and loss rate. The traffic characteristics and real-time nature of these multimedia applications pose new challenges to the design, implementation, and management of future wireless networks [2]. To provide QoS to multimedia applications, a reservation scheme is needed to allocate wireless resources to the different connections. Since the traffic characteristics of most multimedia traffic are in the category of variable-bit-rate (VBR) sources with considerable burstiness, designing an optimal or a good resource allocation scheme becomes especially difficult. If the resources are reserved according to the average traffic rates, unacceptable high delays or unacceptable high packet losses may result when the sources transmit near its peak rate. On the other hand, if the resource reservations are based on the peak rates, the network may be underutilized most of the time [3].

In general, there are two ways to guarantee the QoS of VBR connections: deterministically or statistically. Statistical models employ stochastic traffic models, including Markov modulated, self-similar, S-BIND, among others [6]. These models are either not powerful enough to capture the important burstiness and time correlation of realistic sources, or are too complex for practical implementation. Deterministic models normally characterize their traffic by burst length, average rate, and/or peak rate [3].

In this paper we consider deterministic QoS guarantees in a wireless environment. This is achieved through an accurate traffic characterization of the VBR multimedia traffic streams. We also employ a traffic regulator that can provide bounded packet loss and a traffic scheduler can provide bounded packet delay: Hence, their combination can provide bounded loss ***and*** delay deterministically. This is a distinction from traditional determinitic QoS schemes where a 0% packet loss is always assumed. By allowing a bounded packet loss, it is possible to admit more connections. In particular, most multimedia connections, for instance video and audio, can accept a certain percentage of packet losses while still providing an acceptable QoS.

We performed a set of evaluation experiments. We found that our proposed algorithms can significantly support more connections than a system which do not allow any loss. Moreover, we found that the proposed algorithms out-perform state-of-theart MAC protocols in this area [4], [5], e.g., MASCARA, when the worst-case traffic is being used.

This paper is organized as follows. Section 2 presents our traffic characterization and packet scheduling mechanism. Section 3 and 4 introduce the proposed wireless access and its performance evaluation with our method for providing QoS guarantees. Finally, Section 5 concludes the paper.

2 Traffic Characterization

Before presenting our wireless acces protocol and our QoS guarantee mechanism, we need to accurate chatacterize multimedia traffic. One method that is receiving a lot of attention for characterizing traffic for determinitic QoS guarantee is the (σ, ρ)-model [7]. The (σ, ρ)-model

[1] This research work was supported in part by the Hong Kong Telecom Institute of Information Technology under the grant HKTIIT96/97.EG02.
[2] Email: hamdi@cs.ust.hk, Tel: (852) 2358 6984.

0-7803-5565-2/99/$10.00 © 1999 IEEE

describes its traffic by two parameters: the burstiness σ and an average traffic rate ρ. The traffic constraint function of A (the actual traffic of a connection) is given by:

$$A^*(t) = \sigma + \rho t \quad \text{for all } t \geq 0$$

The model enforces a rate ρ and allows some burstiness up to σ which can be policed with the leaky bucket.

Our proposed traffic model is based on the (σ, ρ)-model but with a small modification. Two more parameters added: maximum traffic rate ρ_{max} and minimum traffic rate ρ_{min}.

The (σ, ρ)-model assumes that all burstiness of traffic arrives at the same time, but in general, the burstiness do not come at the same time and it would normally come within a certain duration. Hence, our objective of adding the maximum traffic rate ρ_{max} is to characterize this fact. Our traffic model becomes similar to the special case of the (σ, ρ)-model with 2 pairs, (σ, ρ) and (0, ρ_{max}).

Most multimedia traffic is continuous,e.g., video and voice connections. Thus the connections have minimum rate. To characterize this fact, we add a minimum traffic rate ρ_{min}. The proposed traffic model has four parameter: burstiness σ, maximum traffic rate ρ_{max}, minimum traffic rate ρ_{min}, and average traffic rate ρ. The traffic constraint function is:

$$A^*(t) = \min(\sigma + \rho t, \rho_{max} t) \quad \text{for all } t \geq 0$$

2.1 Packet Dropping Mechanism

The uniqueness of our QoS mechanism is that we have the capability to bound the loss rate of packets besides bounding their delay. For this reason, we emply a packet dropping mechanism to select the packets to be dropped for each connection. A good packet dropping scheme should satisfy the following criteria: (1) reducing burstiness and rate of the traffic, (2) distributing the dropping of packets evenly, and (3) not exceeding the specified loss rate.

Theorem: The optimal dropping of $L\%$ packets for the worst case of the (σ, ρ, ρ_{max}, ρ_{min}) traffic model is the same as the worst case of the (σ', ρ', ρ'_{max}, ρ'_{min}) traffic model where $\sigma' = \max(0, \sigma(1 - Lf))$, $\rho' = \rho(1 - L)$, $\rho'_{max} = \min(\rho(1 - L), (\rho_{max} - \rho)(1 - Lf) + \rho(1 - L))$, $\rho'_{min} = \rho_{min}$, and $f = \dfrac{1}{1 - \dfrac{\rho_{min}}{\rho}}$.

In the worst case, the traffic of the sources are similar to an extremal periodic on-off model shown in Figure 1.

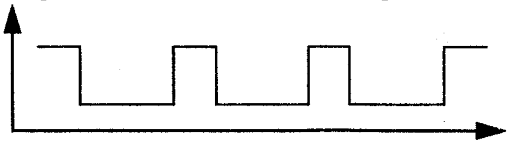

Figure 1: The pattern of worst-case traffic.

An optimal packet dropping algorithm should drop the same amount of packets in every burst. Hence, we can concentrate on only one on-off cycle.

Let:
- ρ_{min} be the minimum traffic rate
- ρ_{max} be the maximum traffic rate
- ρ be the average traffic rate
- σ be the maximum burstiness
- L be the allowable loss rate
- T_{on} be duration of the on cycle
- T_{off} be duration of the off cycle
- T be duration of one on-off cycle

According to the Figure 1, we have:

$$T = T_{on} + T_{off}$$

According to the definition of maximum burstiness, $\sigma + \rho T_{on} \geq \rho_{max} T_{on}$ and $2T_{on}\rho_{max} + T_{off}\rho_{min} \leq \rho(T + T_{on}) + \sigma$.

Since we are considering the worst case, we have:

$$T_{on} = \frac{\sigma}{\rho_{max} - \rho}$$

$$T_{off} = \frac{\sigma}{\rho - \rho_{min}}$$

To maintain desired loss rate L, the maximum number of packets to be dropped is $T\rho L$ in each on-off cycle. Hence,

$$maximum\ allowable\ loss = \rho L\left(\frac{1}{\dfrac{\rho_{max}}{\rho} - 1} + \frac{1}{1 - \dfrac{\rho_{min}}{\rho}}\right)$$

Without violating the desired loss rate, all dropped packets can only be dropped while the traffic is in burst. We find that the resulting traffic can be characterized by the burstiness σ', the average traffic rate ρ', the maximum traffic rate ρ'_{max}, and the minimum traffic rate ρ'_{min}.

As $L\%$ of the packets are dropped, obviously the average rate is also reduced by L. Therefore, we have:

$$\rho' = \rho(1 - L)$$

As packets are only dropped in the burst, the number of packets arrival in time interval $[0, T_{on}]$ is equal to the allowable loss plus the number of packets arrival in the time interval $[0, T_{on}]$ for the new traffic characterization. Therefore, $\sigma + \rho T_{on} = \sigma' + \rho' T_{on} + \sigma L f$ and we have:

$$\sigma' = \max(0, \sigma(1 - Lf))$$

where $f = \dfrac{1}{1 - \rho_{min}/\rho}$.

The length of the ON and the OFF cycle of the resulting traffic is unchanged, so we can find the new value of ρ'_{max} by $\sigma'/T_{on} + \rho'$. Thus, we have:

$$\rho'_{max} = \min(\rho(1 - L), (\rho_{max} - \rho)(1 - Lf) + \rho(1 - L))$$

The minimum rate of the traffic is remained unchanged, so:

$$\rho'_{min} = \rho_{min}$$

According to the Theorem, the traffic of a (σ, ρ, ρ_{max}, ρ_{min})-model with a loss rate $L\%$ can be reduced to (σ', ρ', ρ'_{max}, ρ'_{min})-model in the worst-case scenario of traffic arrival. It is obvious that the traffic of (σ, ρ, ρ_{max}, ρ_{min})-model can be reduced to (σ', ρ', ρ'_{max}, ρ'_{min})-model when traffic arrival is not in the worst case. We can use two leaky buckets to filter the traffic of the (σ, ρ, ρ_{max}, ρ_{min})-model and reduce it to (σ', ρ', ρ'_{max}, ρ'_{min})-model.

2. 2 Packet Scheduling Scheme

Our QoS guarantee scheme is to provide bounded loss and delay. Therefore, we need a packet scheduler once a packet enters the network. The design of the packet dropping scheme and the traffic model are independent of the packet scheduling scheme. Hence, we can use an optimal scheduler, EDF scheduler, as the packet scheduling scheme [3]. Without loss of generality, we assume that the connections are ordered such that $i < j$ whenever $d_i < d_j$ where d is the deadline of packets in a connection. With the proposed traffic model and the packet dropping mechanism, we have the following schedulability condition:

$$t \geq \sum_{i=1}^{j} n_i(\sigma'_i + \rho'_i(t-d_i)) + max_{k<j} s_k^{max}$$

for $d_j \leq t < dj+1,\ 1 \leq j < N,$ and

$$t \geq \sum_{i=1}^{j} n_i(\sigma'_i + \rho'_i(t-d_i)) \qquad for\ (d_N \leq t)$$

As long as the stability condition above is satisfied, for instance, $\sum_{j=1}^{N} \rho_j < 1$, we have:

$$d_j + \frac{\sigma'_j}{\rho'_{max_j} - \rho'_j} \geq \frac{\sum_{i=1}^{j} n_i(\sigma'_i - \rho'_i d_i) + max_{k<j} s_k^{max}}{C - \sum_{i=1}^{j} n_i \rho_i'}$$

3. Medium Access Control Protocols

Recently, many call admission control (CAC) algorithms and medium access control (MAC) protocols have been proposed for multimedia wireless networks [4], [5]. Nearly all of the proposed algorithms only provide statistical or soft QoS guarantees. However, there are many applications that require deterministic, or hard, QoS guarantees. In a wireless network, there are some proposed protocols which can provide deterministic QoS guarantees. But most of these protocols can provide deterministic bounds for delay with zero packet loss [4]. To the best of our knowledge, there are no protocols or algorithms which can provide deterministic bound for BOTH delay and loss guarantees.

Our proposed wireless MAC protocol is motivated by the Bandwidth Reservation Multiple Access (BRMA) [8]. Since the assignment of the mini-slots in BRMA is deterministic, both the request channels and the data channels is contention-free. This characteristic is suitable for us to provide deterministic QoS guarantee. Transmission time is divided into time slots which are further divided into mini-slots and data. Mini-slots are for sending the requests from terminals to the base station, while the data slots are for sending real data or packets to the base station.

In BRMA, each connection will be assigned one mini-slot in each frame. As a result, each connection do not need to content for the channel with other connections. But in our proposed protocol, we have to deal with VBR, CBR, ABR and UBR traffic where CBR and VBR connections need to have deterministic QoS guarantees. In our protocol, only CBR and VBR connections will be assigned one mini-slot in each frame, while ABR and UBR will content for the channel through a random access, e.g., Slotted ALOHA.

4 Performance Evaluation

4.1 Basic Evaluation

We have evaluated the performance of our protocol by varying different factors such as the burstiness σ, the peak rate ρ_{max}, the minimum rate ρ_{min}, and the desired loss rate L. The purpose is to see how these factors affect the performance of the proposed protocol. The results are generated with the parameters given in Table 1 (because of lack of space, we did not draw the results in this paper).

Table 1: Traffic and QoS parameters.

Channel Capacity	45Mbps
Cell Size	53Bytes
ρ	0.15Mbps
ρ_{max}	0.9Mbps
ρ_{min}	0.09Mbps
σ	100 cells
d	30ms
L	5%

First, we investigate the effect of the ratio of the minimum traffic rate to the average traffic rate on the channel utilization. We found that the channel utilization increases from 0.4 to 1 when the ratio of the minimum traffic rate to the average traffic rate increases from 0 to 1. A higher minimum traffic rate implies less variation in traffic; thus the scheduling of traffic is easier and more efficient.

Next, we investigated the effect of the ratio of maximum traffic rate to average traffic rate on the channel ultilization. We found that the channel utilization decreases from 1 to 0.2 when this ratio increases from 0 to 20. In particular, a lower minimum traffic rate implies less variation of the traffic which leads to more efficient scheduling of the traffic.

We studied the effect of traffic burstiness on the channel utilization. We found that the channel utilization decreases from 1 to 0.4 when the burstiness of the traffic increases from 0 to 120 cells. The result showns that lower burstiness of the traffic leads to easier scheduling of the traffic.

Figure 2 illustrates the effect of the packet loss rate on, the channel utilization. The channel utilization increases from 0.368 to 0.425 when the desired bounded loss rate increases from 0% to 5%. By allowing 5% of packet loss for the connections, the channel utilization increases by 15%, thus 15% more connections can be admitted. According to the figure the channel utilization is directly proportional to the desired bounded loss rate.

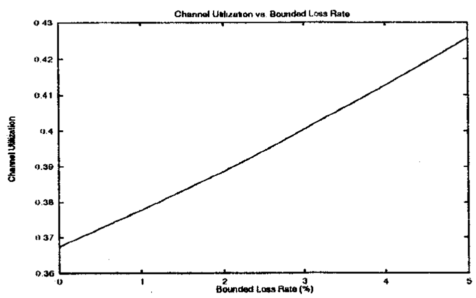

Figure 2: Utilization vs. desired loss rate.

Now, we assume that we have two classes of traffic and the channel capacity is 45Mbps. The other traffic parameters are shown in Table 2.

Table 2: Traffic parameters for multiple types of traffic.

Connection Class	1	2
ρ_i	0.15 Mbps	0.15 Mbps
σ_i	25 cells	250 cells
d_i	30ms	50ms
L_i	5%	5%

According to the Figure 3, the maximum number of class 1 connections that can be supported increases as the number of class 2 connections decreases, and vice versa. When the number of class 2 connections reduces to below a certain value, the maximum number of class 1 connections that can be supported stops to increase. The reason is that the resources saved by not serving class 2 connections are not enough to serve any more class 1 connections as the desired QoS of class 1 connections is much higher than that of class 2 connections. The admission region increases as the ratio of minimum traffic rate to average traffic rate increases. This is due to the fact that the smaller ratio implies less variation of the traffic. When the ratio is 0.8, the area of the admission region of the proposed algorithm is around 40% larger than that of the algorithm which do not allow dropping.

4.2 Comparison with other MAC protocols

To further evaluate the performance of the proposed

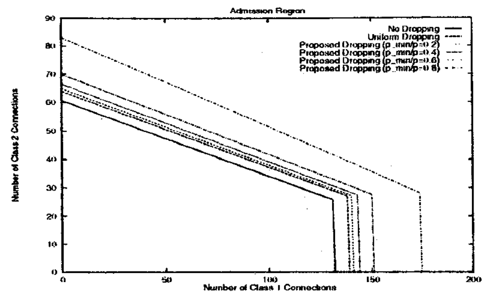

Figure 3: Admission region for two Different classes of traffic.

protocol, we have to compare it with some of the existing MAC protocols. Since there is no wireless MAC protocol which can provide deterministically bounded loss and delay, we are not able to directly compare our deterministic MAC with others. Instead, we compare our proposed protocol with a statistical protocol. We have selected the MASCARA MAC protocol since it is receiving the most attention [9].

As we needed to compare a deterministic MAC protocol with a statistical MAC protocol, we perform the evaluation in two cases, worst and average cases. The high gain from statistical scheduling would favor statistical protocols over deterministic protocols in the average case. In the worst case, the declared QoS of most connections will be violated using a statistical protocol as it does not provide hard QoS guarantees. As a result, we performed the evaluation using three metrics: average loss rate, percentage of connections violating declared QoS, and admission region.

Average Loss Rate

In Figure 4, we found that our proposed protocol outperforms MASCARA in worst case traffic. When the number of users is higher than 56, the average loss rate using MASCARA has exceeded the declared loss rate of 5% in this case. In our proposed protocol, we can still keep the average loss rate below the declared loss rate until the number of users exceed 60.

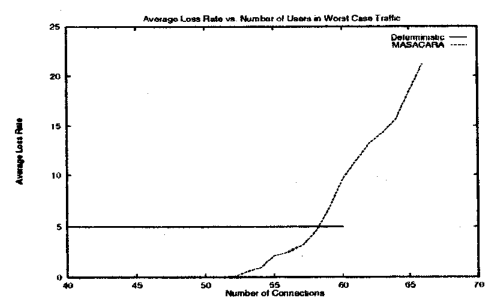

Figure 4: Admission region for two Different classes of traffic.

In Figure 5, we found that our proposed protocol underperforms MASCARA in average case traffic. The average loss rate in MASCARA only exceeds the declared loss rate when the number of users exceeds 75. In our proposed algorithm, we only admit at most 60 connections as it is limited by our CAC algorithm.

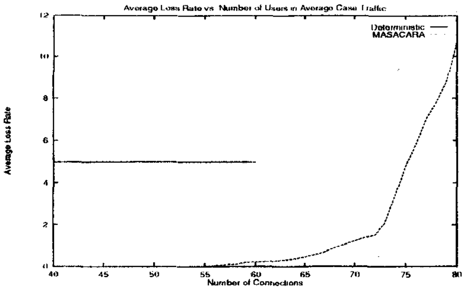

Figure 5: Average loss rate vs. number of users in worse cases traffic.

Percentage of Connections Violate Declared QoS

Sometimes, average loss rate is not a good performance metric. For example, suppose we get only one connection having zero loss while all others exceed the declared loss rate. The average loss rate may still be below the declared loss rate. Thus, we use the percentage of connections violating declared QoS as the metric.

In Figure 6, we found that some connections in MASCARA start to violate the QoS when the number of users increase to 50. The violating percentage increases as the number of users increases. In our proposed protocol, the percentage of connections which violate the declared QoS can be kept to zero when the number of users is less than the calculated value from the call admission control, e.g., 60. When the number of users is 60, there are 23.3% connections having violated the QoS using MASCARA. For the average case traffic, we can find some connections violating the declared QoS only when the number of users is more than 71

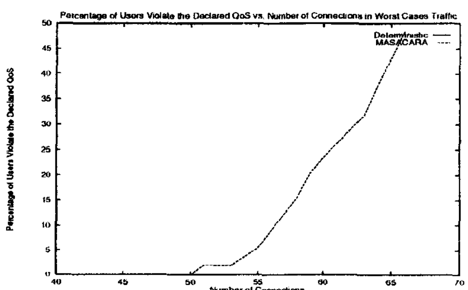

Figure 6: Percentage of connections violate declared QoS vs. number of users in Worse cases traffic.

Admission Region

In Figure 7, we found that the admission region for our proposed protocol is much larger than that for MASCARA in the worst case traffic. In MASCARA, we can only admit 75 type 1 connections or 65 type 2 connections with QoS quarantees. In our proposed protocol, we can admit 100 type 1 connections and 82 type 2 connections under the same condition. On the other hand when we use average case traffic, we found that the admission region for our proposed protocol is less than that for MASCARA. In MASCARA, we can admit 120 type 1 connections or 115 type 2 connections without having any connection violating the declared QoS while our scheme can only admit 105 type 1 connections or 65 type 2 connections.

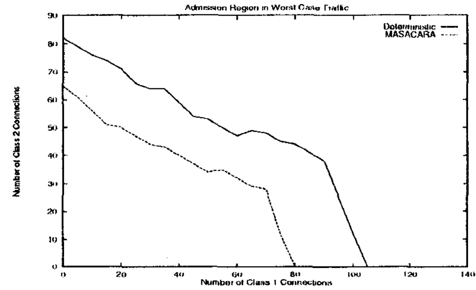

Figure 7: Admission region for worst case traffic.

5 Conclusion

we proposed a *deterministic* QoS guarantees method for wireless networks that accuratly characterizes multimedia traffic streams in conjunction with efficient scheduling and dropping algorithms. The uniqueness of this scheme is that it can provide bounded loss ***and*** bounded delay deterministically. By allowing certain bounded loss for the connections, more connections can be admitted into the network. We have performed a set of performance evaluation tests that demonstrated that our proposed algorithm can significantly support more connections than a system do not allow any loss. Moreover, we found that the proposed algorithms are able to out-perform well-known MAC protocol, e.g. MASCARA, in worst case traffic.

References

[1] A. Acampora, "Wireless ATM: A perspective on issues and prospects," *IEEE Personal Commun.*, pp. 8-17, Aug. 1996.

[2] E. Ayanoglu, K. Y. Emg, and M. J. Karol, "Wireless ATM: Limits, challenges, and proposals," *IEEE Personal Communications*, pp. 18-34, ug. 1996.

[3] J. Liebeherr, D. E. Wrege, and D. Ferrari, "Exact Admission Control in Networks with Bounded Delay Services," *IEEE/ ACM Trans. on Networking*, 1996.

[4] J. Sanchez, R. Martinez, and M. W. Marcellin, "A survey of MAC protocols proposed for wireless ATM," *IEEE Network*, pages 52--62, 1997.

[5] L. Wang, M. Hamdi, and K. B. Letaief, "Wireless channel access for multimedia personal communication," *Proceedings VTC'98*, 1998.

[6] E. W. Knightly, ``Traffic models and admission control for integrated services networks," in *PhD thesis, University of California - Berkeley*, May 1996.

[7] R. L. Cruz, ``A calculus for network delay, part I: Netwrok elements in isolation," *IEEE Trans. on Information Theory*, vol. 37, no. 1, pp. 114-31, 1991.

[8] Z. Zhang and T. Saadawi, ``A bandwidth reservation multiple access protocol for wireless ATM local networks," *Journal of Wireless Networks*, pp. 147-60, 1997.

[9] F. Bauchot, ``Mascara, a mac protocol for wireless ATM," in *Proceedings of ACTS Mobile Summit*, pp. 17-22, 1996.